Nurse rostering under fatigue modelling

June 2019

Abstract

Write abstract last.

1 Introduction

Adverse psychological and physiological effects of night rotations on nurses are well documented [Muecke, 2005]. Impaired vigilance and performance occurs as a result of increased sleepiness and can seriously compromise workers' health and safety [Boivin and Boudreau, 2014], as well as patient safety [Hughes and Rogers, 2004]. This underscores the importance of avoiding the most fatiguing nurse rosters. Different measures are established to facilitate this, most prominently work regulations that aim to hinder employee exhaustion. Such rules and regulations are a key part of the constraints in the Nurse Rostering Problem (NRP) Burke et al. [2004]. Despite their wide application, sleep deprivation and different kinds of fatigue continue to have adverse effects on nurses. In this work, we create a fatigue model by adapting the sleep model from Phillips et al. [2010] to shift work, which provides scores for fatigue levels of nurses. We also create a rolling horizon approximation of the fatigue model, making it useful in Nurse Rostering Problems. Our research demonstrates that the worst cases of fatigue scores can be significantly reduced. Instead of only using traditional shift work rules that deal with short shift transitions, we take into account individual biology in the approximated fatigue model, which further enables finding rosters with low fatigue scores when evaluating the entire planning period. This entails formulating a novel version of the NRP, the Nurse Rostering with Fatigue Minimisation Problem (NRwFMP). **K:** ***Might want to change name? Discuss*** ⇐ Our approach serves as a proof of concept for incorporating a general sleep model in NRP and is generalisable to other rostering and workforce planning problems. To solve the NRwFMP, we develop an algorithm based on a Large Neighbourhood Search (LNS) and present computational experiments. They demonstrate that individuals respond differently to working similar schedules depending on biology, and that managers and planners must take this into account to minimise the risks of fatigue among staff. **K:** ***Update after all computational experiments are final.)*** ⇐

Our main contributions are listed below:

- Creating an approximation of a complex sleep model
- Demonstrating the use of an approximated sleep model in the Nurse Rostering with Fatigue Minimisation Problem (NRwFMP)
- Introducing biological profiles enabling personalised schedules
- Creating new algorithm for solving the NRwFMP
- Managerial insights for minimising the worst cases of fatigue for individual nurses in rosters

The outline of this paper is as follows. In Sections 1.1 and 1.2 we present relevant literature in sleep research and nurse rostering. We include these in a typical NRP in Section 3, and demonstrate how the fatigue model approximation can be utilized despite cases of imprecision. This is done by implementing an algorithm using a Constraint Programming (CP) solver in a Large Neighbourhood Search (LNS) to find high-quality solutions based on the approximation, before verifying the results with the true fatigue model and implementing a post-processing procedure in Section 4. Lastly we provide insights to how different biological characteristics among staff affects rostering with fatigue limitations, and present a trade-off between horizontal and vertical equity in division of night shifts among staff in Section 5. **K:** ***Update description of Section 5 when full experiments are finalized*** ⇐

1.1 Fatigue modeling literature

When models in Operations Research (OR) deal with subjects such as tiredness, stress and work strain, they often present some version of a fatigue model. However, the fatigue term can be ambiguous. Literature in medical sciences and biology often distinguishes between acute fatigue and chronic fatigue, further differentiated into muscular fatigue, mental fatigue, psychomotor fatigue and chronic fatigue associated with post-viral syndromes [Dawson et al., 2011]. In OR, the fatigue term is often loosely defined, if defined at all, and can fully or partly deal with any of the above mentioned variants of fatigue. In Michalos et al. [2010] a job rotation tool designed to provide less monotonous and repetitive tasks for employees is presented. Authors define fatigue as "the physical stress that each process induces on the operators", and it was shown that job rotation plans could reduce the total accumulated physical fatigue per operator [Michalos et al., 2013]. According to Jamshidi [2019], "Fatigue is a stochastic factor that changes according to other factors such as environmental conditions, work type, and work duration", which they handle using chance constraints. In Goel and Vidal [2014], fatigue is not defined explicitly, but rather linked to road transport crashes and falling asleep while driving, in an effort to evaluate regulations. While the different approaches to modelling fatigue in the examples mentioned above are useful, we argue that there is a body of literature within sleep research that models fatigue in a more accurate way when planning work during disadvantageous hours. This literature deals with fatigue fitting the definition provided in Dawson et al. [2011]: "the drive to sleep". We will predominantly use the term fatigue in this work, but note that the term sleep drive can be used interchangeably.

There exist several models of human sleep that can be utilized either directly or as part of quantitative tools to evaluate the fatigue of shift workers, e.g. Borbély [1982], Åkerstedt et al. [2004], Mallis et al. [2004], Hursh et al. [2004], McCauley et al. [2009], Rajdev et al. [2013], St. Hilaire et al. [2016], Postnova et al. [2016]. However, these models tend to be used as tools of retrospective evaluation. It is rare for such tools to be deployed prospectively, i.e. explicitly incorporating them into models that perform planning.

A handful of authors in OR have created or used fatigue models inspired by the body of sleep research literature in proactive fatigue modelling. The fatigue model in Tvaryanas and Miller [2010] is based on the Hursh et al. [2004] "Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE)" model, and incorporates it in a staff scheduling tool, where the SAFTE-model has been used to simulate the fatigue score of all possible schedules. In Wang and Ke [2013], authors were inspired by the Fatigue Audit Inter Dyne (FAID) system [Roach et al., 2004]. Wang and Ke [2013] simplified the FAID model, resulting in a linearization of an exponential function suitable for a MIP framework. The problem considers minimising fatigue in work shift scheduling for air traffic controllers. The linear fatigue model is improved in Wang and Liu [2014], by the addition of a dampening parameter in cases of extreme fatigue. This is shown to fit the results of the FAID model better. A similar model and technique is used for shift scheduling of aircraft maintenance crews in Liu and Wang [2013]. Lin et al. [2013] present a MIP model for nurse scheduling taking into account fatigue using two different approaches. The first approach entails creating a survey for a diverse collection of healthcare facilities in the US. The survey itself was based on the Swedish Occupational Fatigue Inventory (SOFI) [Åhsberg et al., 1997, 2000]. The second approach uses a sinusoidal function that includes a parameter implying individual nurses' chronotype (propensity to sleep at different times), based on work presented in Dawson and Fletcher [2001] to approximate fatigue at the end of a week. The fatigue scores are compared to each other, and interestingly do not correlate. Bowden [2016] proposes the TDSPFM, a Truck Driver Scheduling Model where fatigue is modeled using the non-linear fatigue model proposed in Ingre et al. [2014] model. The Ingre et al. [2014] model is based on the Åkerstedt et al. [2004] three-process model. The non-linear TDSPFM was solved using the evolutionary solution method of the built-in Excel 2013 solver.

The above mentioned OR-models including proactive fatigue modeling are all based on so-called behavioural level models
K: ***Not sure if this is true. Check. Get Andrew's opinion. Otherwise, what are the main differences between these models and our model?*** ⇐

FAID has some flaws, but do we have citeable material showing this? Quite sure the model we are using is more advanced. Is it fair to call it state of the art?
that aim to describe key concepts of sleep-wake behavior [Postnova, 2019]. Although the behavioural level two-process model of Borbély [1982] has served as a conceptual framework in sleep research recent decades, and is considered the most used mathematical model for sleep, significant advances have been made in the wake of it, including models at the level of brain areas [Postnova, 2019]. Examples of such models are Rajdev et al. [2013], Ingre et al. [2014], St. Hilaire et al. [2016], Postnova et al. [2016] and Phillips et al. [2010].

The Phillips et al. [2010] model combines the [Phillips and Robinson, 2007] model of the ascending arousal system with the Forger et al. [1999] human circadian pacemaker model. The model is written in matlab and is solved using a built-in ordinary differential equation solver. It outputs fatigue levels and sleep/wake states for the time period evaluated. The Phillips et al. [2010] model has been subject to testing and parameter-tuning, and similar models of have been based on it since. **K:** ***Are there other articles published on this? In that case, should be referenced.*** It is one of the most advanced and precise sleep models currently in existence, and we have gotten full access to it for this work. Thus it is used as the cornerstone of the fatigue model developed in Section 2. ⇐

⇐

K: ***The paragraph above is designated to introduce and shortly describe Andrew's sleep model. Hoping Andrew can make improvements to it. I am not sure if the functionality of forced wakefulness was in the original sleep model in the paper from 2010. Does not look like that to me. Therefore, I mention this functionality in section 2 about the fatigue model rather than here. If this is incorrect, or it is added in some other paper previous to our project, it would be great if Andrew could edit.***
While we aim to incorporate a complex sleep model in Nurse Rostering in this work, the intricacies of the Phillips et al. [2010] model will not be discussed in this work. Rather, we note that for our use, external factors that can affect sleep, but that are not directly linked to rostering decisions, are left fixed at default values or not modelled. Examples of such factors are times of dawn and dusk or a person's intake of caffeine. Furthermore, the output that is most relevant for our use is the sleep drive (fatigue), which typically exists in an interval [-2mV,8mV] depending on biological parameters and factors that affect sleep.

1.2 Nurse Rostering literature

The NRP is a scheduling problem dealing with assigning a number of shifts with predefined start and end times to a set of nurses in a given planning period. NRPs typically include coverage constraints, i.e. some constraints ensuring a minimum number of nurses on duty, time related constraints e.g. a number of hours to be worked during the planning period and a set of work regulations [Burke et al., 2004]. A range of different rules and regulations exist in Nurse Rostering literature. The many different variations are too many to mention explicitly, but for additional details, we refer readers to Burke et al. [2004] and [Haspeslagh et al., 2014]. As no widely accepted standard NRP exists, we create an NRP based on guidelines from Safe Work Australia [Safe Work Australia, 2013].

NRPs are solved in numerous ways, e.g. Artificial Intelligence (AI) approaches, CP, metaheuristics and mathematical programming approaches [Ernst et al., 2004]. In the realm of CP, Downing [2016] tackle Nurse Rostering, among other problems, with lazy clause generation. Other examples of CP include Pizarro et al. [2011] and Métivier et al. [2009]. Examples of AI methods include Meyer auf'm Hofe [2001], which builds on CP and integrates fuzzy constraints with branch and bound Musliu et al. [2002]. The hybrid artificial bee colony algorithm presented in Awadallah et al. [2015] is another AI method used, where the bee operator is replaced with the hill climbing optimizer. According to Burke et al. [2008], metaheuristic methods seem to be the dominant technique when solving real-world problems. Examples are the tabu-search based metaheuristic of Rönnberg and Larsson [2010] and the case-based reasoning approach of Beddoe et al. [2009].

In the realm of mathematical approaches, the standard mixed integer programming models are among the most explored ones, see e.g. Ásgeirsson and Sigurardóttir [2016] and Mischek and Musliu [2019]. Different decomposition methods have also been explored, with variants of column generation and branch and price implementation being popular modeling choices [Dohn and Mason, 2013], [Beliën and Demeulemeester, 2007].

The approach to handling fatigue constraints introduced in this paper have been integrated into a CP implementation of nurse rostering. We believe it can also be combined with mathematical approaches. Integration with heuristic methods, such as bee colony optimisation and case-based reasoning, would require a redesign of the heuristics.

Notably, literature on Nurse Rostering often focuses on solution techniques [Petrovic and Berghe, 2012]. We argue there should be an increased focus on creating models that are useful in practise and that provide insights for real-life decision makers. We aim to do so by creating a tool producing fatigue minimising rosters that reduce risks of accidents and improve nurse health, and create managerial insights for decision makers based on our computational results.

2 The fatigue model

In this section we present the fatigue model based on the sleep model of Phillips et al. [2010]. The sleep model includes a sleep/wake switch, which models how a human falls asleep and wakes up as a result of internal processes in the brain and light conditions. The impact of shift work on our model is that it precludes sleep. This implies that the times a person is at work, the fatigue model is restricted from entering a sleeping state. We have defined four shifts in accordance with Safe Work Australia [2013] guidelines for managing the risk of fatigue at work to represent realistic and advisable shift times:

- Day shift “D” 07:00 - 15:00
- Evening shift “E” 14:30 - 22:30
- Night shift “N” 22:00 - 07:30(+1 day)
- Off-shift “O”

To add to the realism, we have chosen to include 45 minutes of forced wake-time before and after work in our fatigue model, to represent commuting. Depending on the roster a nurse works, the fatigue model calculates fatigue based on his or her shifts.

The model has some default starting values for its variables. The initial values of the fatigue model variables reflect a well-rested individual where the circadian rhythm has been given time to stabilize in the individual’s preferred phase. To ensure this, we simply let the sleep model run for long periods without any work, thus obtaining the default initial fatigue model state.

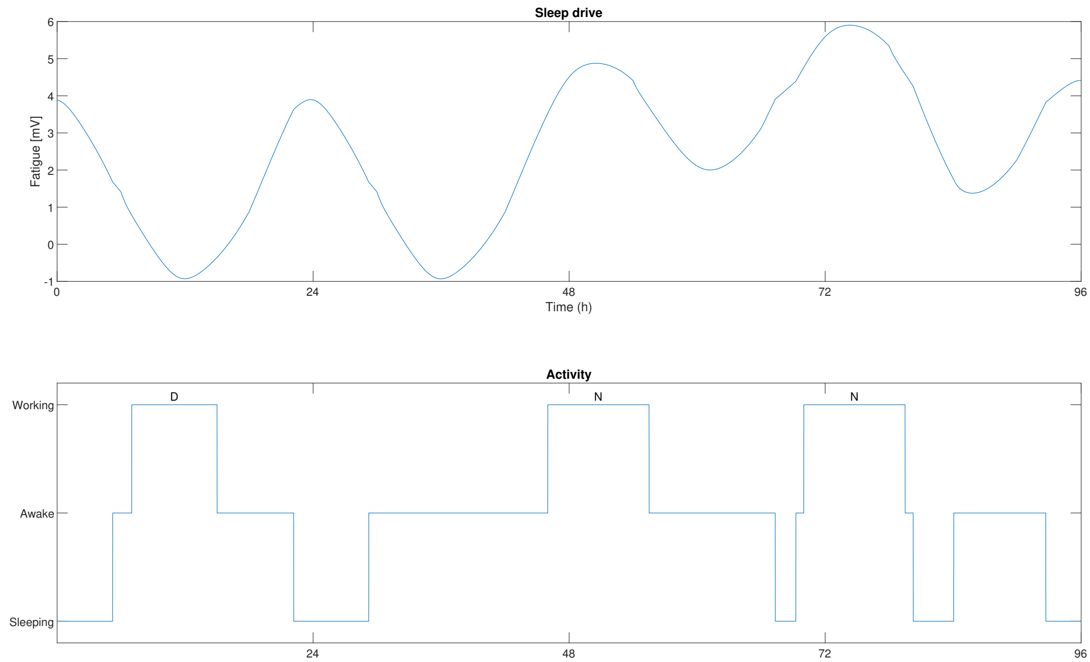


Figure 1: Plots of how fatigue and activity of a nurse with typical biological parameters change as time passes. The nurse is scheduled to work the 4-day roster {D,N,N,O}.

In Figure 3, two plots of a four day example roster {'D','N','N','O'} for a nurse of a typical biological profile is presented. The fatigue is represented in the top plot and clearly changes through time each day as a result of the person's circadian phase. Furthermore, as the person works night shifts, the fatigue increases. In the bottom plot, the activities at all times in the form of sleep, wakefulness and work are presented. Note that night shifts are defined to begin during the end of a day, so the night shifts during days 2 and 3 begin at hours 46 and 70. From the activity plot it is clear that the nurse only got a short period of sleep between night shifts, falling asleep around hour 67, before the nurse was forced to awaken 45 minutes prior to the second night shift beginning at hour 70.

In this work, the fatigue model is regarded as the best representation available of a nurse's true fatigue at any time, and the fatigue scores provided by the model are thus sometimes referred to as the true fatigue of a nurse. In Section 2.2 the fatigue model will be approximated for incorporation in NRPs. It will be referred to as the rolling horizon approximation or simply the approximation, and fatigue scores obtained using it will be referred to as the approximated fatigue of a nurse.

2.1 Biological profiles

In this work we wish to take into account that fatigue develops differently for different individuals. Two parameters represent the most common differences in biology related to sleep. **K: *Need Andrew's input for citations here*** ⇐ These parameters affect the normal sleep-time and the chronotype of a human.

The average sleeptimes-parameter is calibrated so that a person that is not working has an average sleep time of the most common ≈ 7 hours, short sleep time ≈ 5 hours or long sleep time ≈ 9 hours. For chronotype, the parameter is either set to its standard value representing the most common "day-time chronotype", early "morning-type" or late "evening-type". We thus get 9 different biological profiles, which we perform analyses on in Section 4. The most typical biological profile was used when plotting the fatigue in Figure 3.

2.2 Approximating the fatigue model

The fatigue model is inherently non-linear, and incorporating it in an NRP is not trivial. When creating a parameter to represent the fatigue model in our NRP, time should be discretisised into days. We believe the most relevant value to represent a nurse's fatigue throughout a day is the highest fatigue experienced during the 24 hours of that day. Evaluating the fatigue created by all possible rosters of realistic sizes is unrealistic. The number of possible rosters is simply too large. For example, for a roster with 4 shifts and a planning period of 42 days an upper bound for the number of possible rosters would be 4^{42}. The number of feasible rosters would be lower, but for realistic sets of constraints, the number of rosters is still huge.

To deal with this issue, we develop a rolling horizon approximation of the fatigue model, using explicit enumeration of all possible rosters of a given number of days *hor* (for *horizon*). The rosters of length *hor* are stored in a lookup table. This approach implicitly assumes there exists a finite number of days (hor) shorter than the planning horizon of the full roster, that we can evaluate to get a useful approximation of the fatigue, when running the model with a default initial fatigue model state. When evaluating a time period $[t - hor + 1, t]$, this period is referred to as the evaluation horizon. The sequence of shifts worked during the evaluation horizon is referred to as the evaluation pattern. For every evaluation pattern we elicit a measure of the nurse's fatigue during the final shift on day t.

Clearly the longer the horizon, the better the estimate. The best estimate from the model is, of course, when the complete work history of the nurse is entered into it: in effect this is an infinite horizon. The action of performing an

Day	1	2	3	4	5	6	7
indrost	N	N	O	D	D	E	O
indpat3	**N**	**N**	**O**				
indpat4		N	O	**D**			
indpat5			O	D	**D**		
indpat6				D	D	**E**	
indpat7					D	E	**O**

Table 1: Demonstration of how the rolling horizon approximation evaluates the different 3-day patterns that exist as parts of indrost. The rolling horizon approximation uses the information from the last day of the evaluation patterns, and save them to comprise the approximated fatigue scores for all days. Shift codes in bold represent the scores stored for each evaluation pattern.

evaluation of a full individual roster in its entirety, thus obtaining the model's best possible prediction of the fatigue scores (the true fatigue), is referred to as a Full Roster Evaluation (FRE). The action of performing an evaluation of an individual roster using the rolling horizon approximation is referred to as a Rolling Horizon Evaluation (RHE).

For example, suppose the 7-day individual roster $indrost = \{N, N, O, D, D, E, O\}$ begins on day 1 and that it is approximated using a 3-day rolling horizon approximation. We wish to record the nurse's fatigue state on each day of the shift. The best fatigue prediction will evaluate the full *indrost* and store the fatigue scores each day, while the approximation will evaluate the 7 different 3-day individual rosters $indpat3..indpat7$ and store the fatigue score obtained on the last day, as demonstrated in Table 1. For days 1 and 2, the daily fatigue scores obtained in $indpat3$ are used. Obviously, the 3-day RHE results will match the FRE results for days 1..3 due to identical initial values, but from thereon differences may arise.

The default initial fatigue model state, before an RHE (or FRE), was introduced above. In fact note that every evaluation pattern can follow a night shift or a non-night shift. Shifts worked prior to our evaluation horizon should not have an impact on our rolling horizon approximation, but as night shifts stretch into the following day and forces a state of wakefulness in the beginning of the first day of the evaluation horizon, this information must be taken into account. For this reason, we introduce an additional initial fatigue model state for all evaluation patterns for each biological profile, which should be used whenever an evaluation pattern in a RHE succeeds a night shift. This additional initial fatigue model state is found for each biological profile by running our sleep model for a one-day roster $\{N\}$ from the default initial fatigue model state, and saving the parameter values. See for example $indpat5 = \{O, D, D\}$ in Table 1. A night shift is worked prior to the first day of $indpat5$, which impacts the approximation. When performing the 3-day RHE, $indpat5$ should thus be evaluated from the additional initial fatigue model state beginning after a night shift.

In the NRP, the fatigue is represented by the parameter $fscore_{b,w,epat}$ for any nurse of biological profile b, following a night shift if $w = 1$ and not if $w = 0$, working evaluation pattern $epat$. The fatigue score is represented as the parameter $fscore_{b,w,epat}$.

2.3 Testing the rolling horizon approximation

We run our FRE and our RHEs for different evaluation horizons on a collection of 30 real-life rosters of 42 days worked by anonymous nurses at the Austin hospital in Melbourne to evaluate the quality of our approximation. We perform our analysis with $hor \in [3..7]$, as preliminary testing implies $hor \leq 2$ is insufficient and $hor \geq 8$ gives an impractically large parameter for the fatigue score both in terms of enumerating all rosters and in terms of complexity in our NRP in Section 3. The first hor days of the RHEs will naturally be identical to the FRE.[1] Thus, we disregard the data for the first 7 days. This provides us with 30 rosters of 35 days for 9 biological profiles. Every day, in each roster, for all biological profiles, we identify the maximum fatigue values, and thus get 9450 data points to compare the FRE with each of the RHEs.

A difference in sleep drive of 0.1mV is regarded as insignificant by the developer of the Phillips et al. [2010] model, thus there is a good match if in most cases the errors between FRE and RHEs are less than this. Performing an FRE entails solving a differential equation using the MATLAB ordinary differential equation solver "ode23", see Shampine and Reichelt [1997]. The 0.1mV benchmark for magnitudes of errors necessitated a significant reduction in the tolerances of the differential equation solver, as compared to the default values.

To evaluate the quality of the rolling horizon approximations of different evaluation horizons, we want to compare each data point in the FRE with each data point in the RHEs, by quantifying the errors of the approximations. For every RHE of a given evaluation horizon, for each data point, we subtract the value provided by the RHE from the value provided by the FRE for the same data point. E.g., for a 3-day rolling horizon approximation we find the value of $FRE - RHE_3$ for all 9450 data points. We then sort the errors, and obtain percentiles to get an overview of how large the errors are.

[1] except for possible errors due to the use of numerical methods in the differential equation solver

Evaluation Horizon	1st perc.	5th perc.	10th perc.	90th perc.	95th perc.	99th perc.
$FRE - RHE_3$	-1.5024	-0.3074	-0.1169	0.0224	0.0895	0.7230
$FRE - RHE_4$	-1.3593	-0.3014	-0.0969	0.0144	0.0631	0.6078
$FRE - RHE_5$	-1.1767	-0.2859	-0.0879	0.0088	0.0454	0.5438
$FRE - RHE_6$	-1.0473	-0.2606	-0.0743	0.0058	0.0411	0.4636
$FRE - RHE_7$	-1.0355	-0.2019	-0.0551	0.0049	0.0389	0.4968

Table 2: Results of subtracting values of RHEs of different evaluation horizons from the FRE of 30 real rosters. Values for all biological profiles are used. All units in millivolts.

Results of the analysis are presented in Table 2. Firstly, we note that errors decrease for longer evaluation horizons, as one would expect. For all percentiles, the longer evaluation horizons have errors closer to 0 in Table 2. Secondly, it is notable how the magnitude of the errors are larger than the insignificant magnitude 0.1 in several cases (less than -0.1 or more than 0.1).

To understand how errors occur, we present a figure consisting of two plots of one of the real 42-day rosters evaluated. The roster is as follows:
O O O N N N N O O O N N N N O D O N N N O O O O N N N N O O O N N N N O O O N N N O
This roster would be considered challenging for most nurses, and is in violation with many of Safe Work Australia's guidelines, but is well suited to demonstrate how errors in our approximation can occur. Consider the one-week excerpt of a 42-day schedule presented in Figure **??** (a similar figure of the full 42 days are provided in Appendix A).

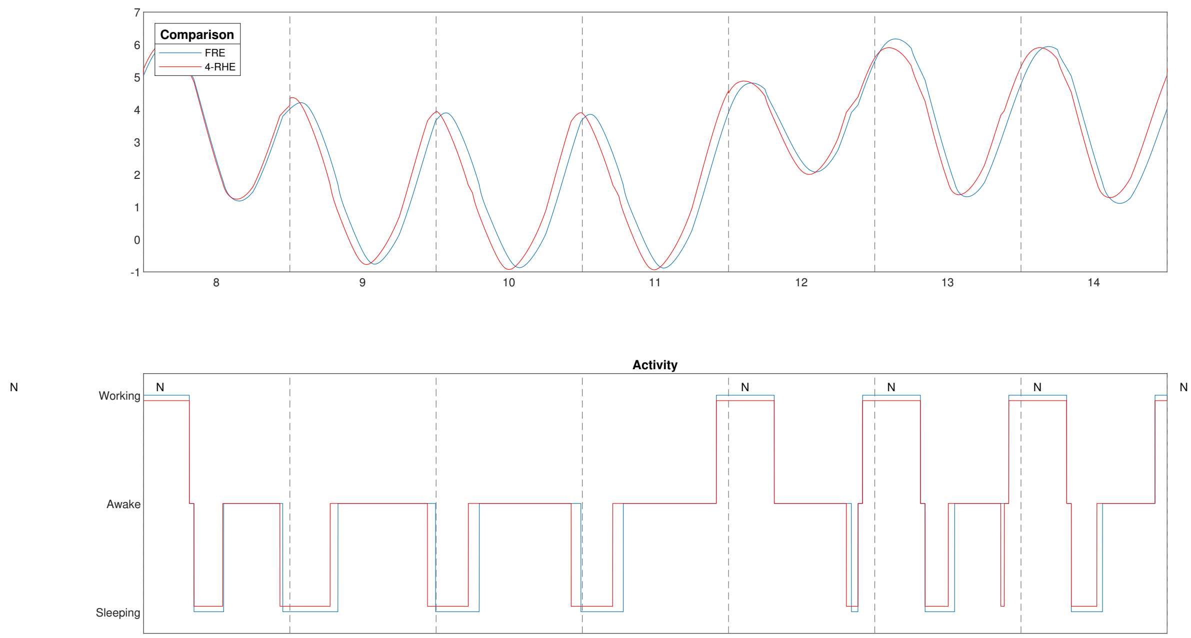
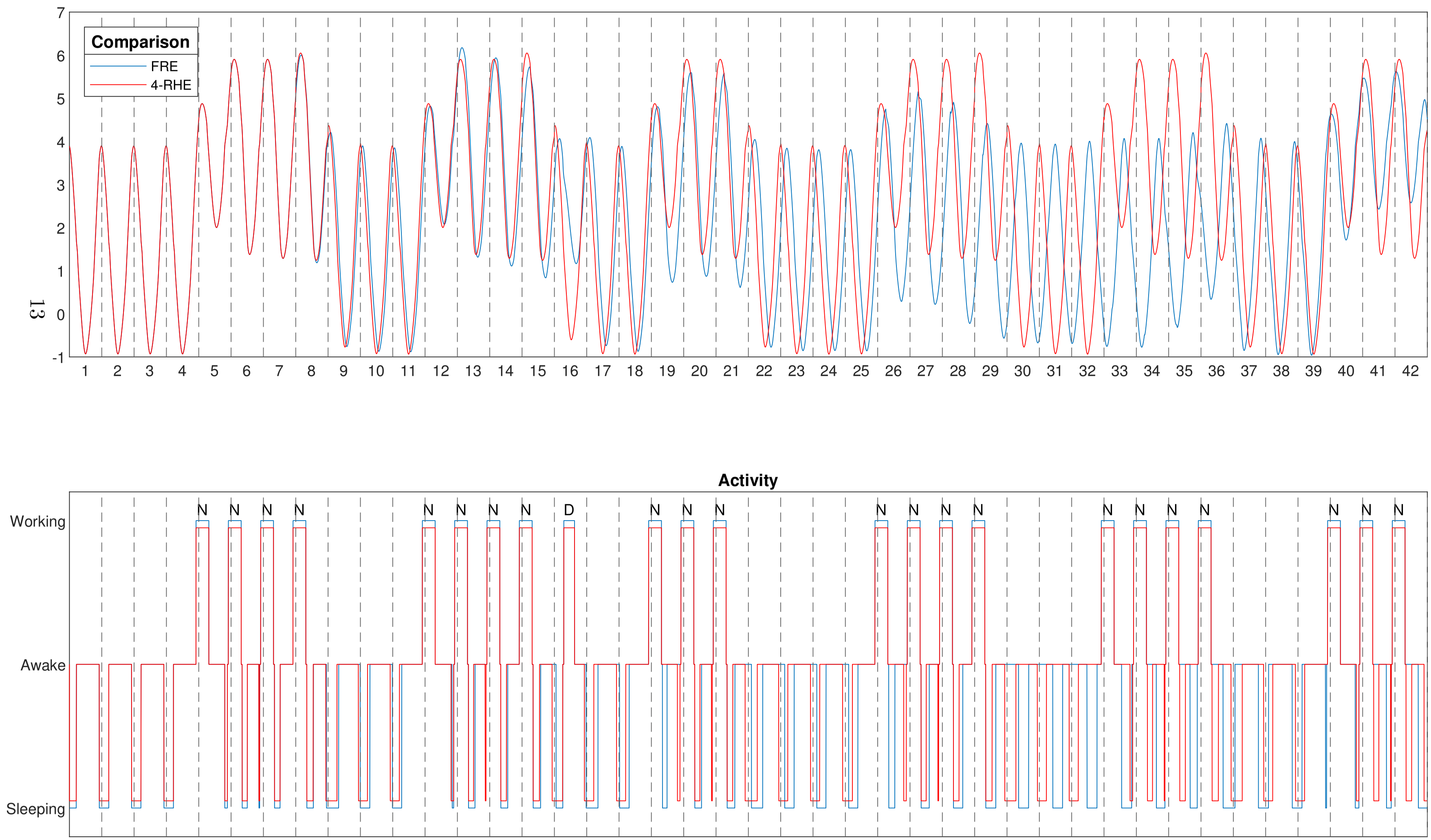


Figure 2: Excerpt of the second week of our exemplary roster.

In Figure **??**, a FRE is compared to a 4-day RHE, with the fatigue of a nurse of biological profile 1 is presented in the top plot, while the nurse's activity as modelled by the FRE and RHE_4 is presented in the bottom plot. During the first 8 days of the roster, the FRE- and RHE-fatigue is aligned and errors are very small. However, at the beginning of day 9 the two graphs have a gap between them. This gap is transferred to day 10, 11, etc. This is typical for the patterns we have evaluated with significant errors. The FRE has recognised that the nurse's circadian rythm has shifted, and graphs have similar shapes, but are unphased.
This unphasing in the FRE does not necessarily imply large errors when comparing the maximum daily values with the maximum daily values of RHEs. However, the unphasing can lead to the RHE assuming a different sleep pattern than the FRE, which again can lead to huge errors in later parts of the roster. Notice the bottom plot in Figure **??**. On day 13, the RHE assumes the nurse will have a short period of sleep before the night shift begins late that day. And as the Figure in the Appendix demonstrates, the RHE remains erroneous throughout the roster.
Put in Appendix:


Comparison
FRE
4-RHE
Activity
Working
Awake
Sleeping
N
D

This highlights the main weakness of the design of our approximation. Due to the limited time horizon, we do not have the same history available as in the real FRE case.

3 the Nurse Rostering with Fatigue Minimisation Problem

The following hard constraints are based on Safe Work Australia's guide for managing the risk of fatigue at work Safe Work Australia [2013]. The number of successive night shifts are restricted to 3, and there is an upper limit to the number of hours worked in any week of 50 hours. Nurses have the weekend off at least every third week. Forward rotation ensures rest between shifts. Two consecutive off-days should be ensured for every nurse with a reasonable frequency. After ending a night shift, or a sequence of consecutive night shifts, every nurse should have two consecutive nights without work. Work weeks are limited to maximum 6 shifts from Monday to
The Nurse Rostering model is formulated as a base model including only the most essential constraints. Furthermore, we define three extensions to the model, including constraints and objectives related to fatigue, safe work rules and preferences. For the purposes of our experiments nurses are identical up to their biological profiles. Thus, we avoid any form of history prior to the first day, fixing all shifts to off-days before day 1. NRP in MiniZinc [Nethercote et al., 2007] and

3.1 The base model

The CP problem contains the following parameters and variables.

Sets Ns = 1..30 - set of nurses
T = 1..42 - set of days
Sh = day,eve,ngt,off - set of shifts

Parameters $cov_s : s \in Sh$ - coverage needed on shift s, day t
$h_s : s \in Sh$ - length in hours of shift s
$hmax_n : n \in Ns$ - maximum work hours for nurse n over the planning horizon

Decision Variables $On_{n,s,t} : n \in Ns, s \in Sh, t \in$ T - 1 if nurse n works on shift s on day t

Constraints

$$\sum_{n \in Ns} On_{n,s,t} \geq cov_{s,t}, \forall s \in Sh, t \in \mathrm{T} \tag{1}$$

$$\sum_{t \geq 0 \in \mathrm{T}} \sum_{s \in Sh} h_s * On_{n,s,t} \leq hmax_n, \forall n \in Ns \tag{2}$$

$$Objective = satisfy \tag{3}$$

Constraints (1) ensure coverage of the demand for nurses. Constraints (2) ensure nurses do not exceed their contracted maximum number of work hours. Unless an extension that includes an objective function is added to the base model, it is solved as a satisfaction problem, where any feasible solution is accepted.

3.2 Extension: Shift work rules

The shift work rules-extension adds the following sets, parameters, variables and constraints to the problem.

⇐

K: *Must be updated with globals*

Sets $Sundays = 7, 14..42$ - set of days

Parameters $wmax = 50$ - maximum working hours in any week

Decision Variables $Tw_{n,t} : n \in Ns, t \in \mathrm{T}$ - 1 if nurse n has two days off ending on day t

Constraints

$$\sum_{d \in t-3..t} On_{n,ngt,d} \leq 2, \forall t \geq 0 \in \mathrm{T} \tag{4}$$

K: *Was this implemented as written here? the t-index will get lower than 0, did we define days 2 weeks prior to off like discussed?* ⇐

$$\sum_{d \in t-6..t} \sum_{s \neq off \in Sh} h_s * On_{n,s,t} \leq wmax, \forall t \in Sundays, n \in Ns \tag{5}$$

$$\begin{aligned}
&(Tw_{n,t} = 1 \leftrightarrow \\
&\qquad On_{n,ngt,t-2} = 0 \wedge On_{n,off,t-1} = 1 \wedge On_{n,day,t} + On_{n,eve,t} = 0 \\
&\quad), \\
&\forall t \geq 0 \in \mathrm{T}, n \in Ns
\end{aligned} \tag{6}$$

$$Tw_{n,t-14} + Tw_{n,t-7} + Tw_{n,t} \geq 1, \forall t \in Sundays, n \in Ns \tag{7}$$

$$\sum_{d \in t-9..t} Tw_{n,d} \geq 1, \forall t \geq 0 \in \mathrm{T}, n \in Ns \tag{8}$$

Constraints (4) ensure no nurse works more than 3 consecutive nights, while Constraints (5) restricts working more than $wmax$ hours every week. Constraints (6) ensure that the variable $Tw_{n,d}$ is 1 if nurse n is off work for two consecutive days ending on day d. Constraints (7) ensure that all nurses have at least one weekend off every 3 weekends. Furthermore, every nurses should have two consecutive days off at least once every 10 days, as instructed through Constraints (8). Note that we define an off-weekend as not working night shifts Friday nor Saturday and not working any other shifts on Saturday nor Sunday.

$$On_{n,ngt,t-1} + On_{n,day,t} + On_{n,eve,t} \leq 1, \forall t \in \mathrm{T}, n \in Ns \tag{9}$$

⇐

MARK: $On_{n,eve,t-1} + On_{n,day,t} \leq 1$ ***??***

$$On_{n,eve,t-1} + On_{n,day,t} \leq 0, \forall t \in \mathrm{T}, n \in Ns \tag{10}$$

$$On_{n,ngt,t-2} + On_{n,off,t-1} + On_{n,ngt,t} \leq 2, \forall t \in \mathrm{T}, n \in Ns \tag{11}$$

$$\sum_{d \in t-6..t} \sum_{s \in day,eve,ngt} On_{n,t,s} \leq 6, \forall t \in \mathrm{T}, n \in Ns \tag{12}$$

Constraints (9) and (10) make sure that forward rotation is not possible, thus securing a minimum period of rest between shifts for all nurses. Constraints (11) ensure that it is not possible to work a night shift followed by an off day and then another night shift, while (12) sets the maximum number of consecutive work days to 6.

3.3 Extension: Fatigue model

⇐

MARK: ***The fatigue model below is hard to understand - I think it needs illustrating with a running example. The example itself should be introduced informally before the formal fatigue model is presented.***
K: ***I like the thought of that. Will try to come up with one.*** ⇐
The following sets, parameters and constraints are added to the base model.

Sets
$Bios := 1..9$ - set of biological profiles
$Ns_b := 1..10$ - set of nurses, each with a biological profile b
Parameters
hor is the size of the evaluation horizon in the rolling horizon approximation of the fatigue model. The evaluation horizon will thus have a range $[t - hor + 1, t]$ at any day t.
P^{Pat} denotes any evaluation pattern of a length corresponding to the evaluation horizon hor.
$epat_{bwP^{Pat}}$ is the highest fatigue obtained during the last day of the evaluation pattern P^{Pat} for a nurse of biological profile b, with $w = 1$ or $w = 0$ indicating

a night shift or not before the first day of P^{Pat}.

Decision variables f^{Max} is the highest fatigue value occurring for any nurse at any time during the planning period $y^{Pat}_{nwP^{Pat}t} \in \{0..1\}$ is a binary variable determining if shift pattern P^{Pat} is worked by nurse $n \in \mathcal{N}$ on day $t \in \mathcal{T}$. If $w = 1$ it indicates the shift pattern follows a night shift, 0 else.

$\Leftarrow$

MARK: y_{nts} ***- what is the*** s ***there for?***
We also define the use of the operator $\Rightarrow$ so that if $y_{nts} \Rightarrow P^{Pat}$, nurse n works shift pattern P^{Pat} ending on day t.

Constraints

$$epat_{bwP^{Pat}} \leq f^{Max} \iff y_{nts} \Rightarrow P^{Pat} \wedge y_{n('N')(t-hor)} = w,$$
$$n \in \mathcal{N}^B_b, w \in \{0,1\}, t \in \mathcal{T}, s \in \mathcal{S}, b \in \mathcal{B}, P^{Pat} = \{s_{t-hor+1}, .., s_t\} \tag{13}$$
$$\text{Objective} = minf^{Max} \tag{14}$$

K: ***Add constraint for circadian rhythm*** $\Leftarrow$
Constraints (13) ensure that f^{Max} is as high as the most fatiguing shift pattern worked by any nurse at any time in the planning horizon, while Constraint (14) minimizes that fatigue score.

4 Computational Study

The Nurse Rostering model and its extensions are modeled using the MiniZinc IDE, which allows for a choice of different CP and MIP-solvers. Preliminary testing demonstrates that no high-quality solution is found within reasonable time using exact methods, and we develop the algorithm presented below. The algorithm is run using Python3.6.8 and Matlab R2018 is used when performing FREs in post-processing.

1. **Find feasible solution.** Run the base NRP from Section 3.1 in MiniZinc using built-in solver Gurobi 7.0.2.

2. **Reduce the highest rolling horizon approximated fatigue score until it is below the threshold.**
We define a neighborhood by fixing the shifts found in Step 1 for all nurses except the nurse with the highest rolling horizon approximated fatigue score and a set of randomly chosen additional nurses. We also do not fix any off-shifts. We then solve the NRP by minimizing the highest rolling horizon approximated fatigue score. If it is unchanged and observed at the same day for the same nurse as in the previous iteration, increase the maximum time limit for the iteration.

3. **Perform post-processing.** Choose the best solution from Step **??**, and perform our simple post-processing procedure: check fatigue scores of the solution using the real fatigue model. Find the highest real fatigue value for any nurse on any day. For the nurse and day with the highest fatigue observed, force an off-day shift the day before **MARK: *Why?*** ⇐ and reoptimize based on approximated values using the LNS described in Step **?? MARK: approxstep*?*** ⇐ to obtain a new roster. Evaluate the real fatigue values of the new roster. If the highest real fatigue score of the new roster is lower than the roster found in Step **??**, the new roster is considered the current best solution and Step 3 is repeated. Otherwise, the current best solution remains unchanged and we continue to Step 4.

4. **Acceptance.** Accept current best solution and terminate search.

Tests are performed using a **K: *Insert computer specs...*** ⇐

4.1 Instances

To ensure an informed basis for analysis, we wish to perform experiments on a realistic set of instances. Our biological profiles are designed to replicate how typical (biological profile 1) adults are affected by shift work, while the probability of having short sleep time or long sleep time and being morning-type or evening-type are all considered $\approx$ 10% of the population. **K: *(Ask Andrew, but I believe this is what he told me)*** ⇐ Thus, to create realistic instances, we draw the two characteristics that constitute the biological profile independently to create a set of n instances. The probabilities are presented in Table 4.

		Average sleeptimes		
		Normal (∼7h)	Short (∼5h)	Long(∼9h)
Chronotype	Day-type sleeps ∼22	1	2	3
	Morning-type sleeps ∼21	4	5	6
	Evening-type sleeps ∼24	7	8	9

Table 3: The 9 biological profiles.

Table 4: Illustration of the probability of drawing different biological profiles for the nurses in our instances. The two characteristics constituting the default biological profile are normal sleep time and having the day-type chronotype. As both characteristics have a probability of 0.8 of being a randomly chosen nurse's characteristic, the total probability of belonging to the default biological type is 0.64. The same logic applies to drawing all the biological profiles.

			Normal sleep time		
			Short ≈ 5	Normal ≈ 7	Long ≈ 9
		Probabilities	0.1	0.8	0.1
Chronotype	Morning-type	0.1	0.01	0.08	0.01
	Day-type	0.8	0.08	0.64	0.08
	Evening-type	0.1	0.01	0.08	0.01

4.2 Preliminary experiments for typical biological types

Table 5

		Fatigue threshold						
		540	550	560	570	580	590	600
Instance 1	RH approx. max fatigue score	488	590	544	488	553	490	544
	RH approx. avg. of ind. max fatigue scores	442	548	438	429	472	426	426
	True max fatigue score after PP	488	594	548	488	572	490	599
	True avg. of ind. of max fatigue scores after PP	449	553	454	445	480	442	540
Instance 2	RH approx. max fatigue score	492	487	544	544	544	544	544
	RH approx. avg. of ind. max fatigue scores	438	432	434	436	430	428	424
	True max fatigue score after PP	504	489	546	549	550	553	543
	True avg. of ind. of max fatigue scores after PP	462	436	449	452	445	448	446

4.3 Neighbourhood analysis

$\Leftarrow$

MARK: ***We previously specified a particular neighbourhood (worst fatigue). This is the point where it should be described.***
To conclude on a reasonable neighbourhood we present tests of proposed neighbourhoods in our LNS, and decide on the preferred one for further testing.

Table 7: Results of running n instances for each case.

Instances	Random neighbourhood	Worst fatigue	Combo
1 BM + FE	769	811	811
2 BM + FE	658	658	658
3 BM + FE	707	714	714

Table 6

		Fatigue threshold								
		520	530	540	550	560	570	580	590	600
Instance 1	RH approx. max fatigue score	506	506	535	554	535	564	564	590	597
	RH approx. avg. of ind. max fatigue scores	470	471	488	484	484	490	484	548	530
	True max fatigue score after PP	601	507	536	601	601	601	601	594	599
	True avg. of ind. of max fatigue scores after PP	539	473	492	539	539	539	539	549	557
Instance 2	RH approx. max fatigue score	516	516	553	553	553	564	564	590	599
	RH approx. avg. of ind. max fatigue scores	468	482	509	505	470	480	493	545	504
	True max fatigue score after PP	516	659	564	562	558	589	574	592	605
	True avg. of ind. of max fatigue scores after PP	472	545	510	508	485	492	508	538	508

Table 8: Results of running n instances for each case.

Instance	Case	Run time	Approx. Obj. Val.	Real Obj. Val.	Optimality gap
1	BM + SR		NA	652	
	BM + FE		656	661	
	BM + SR + FE		651	659	
2	BM + SR		NA	662	
	BM + FE		599	600	
	BM + SR + FE		634	658	
3	BM + SR		NA	601	
	BM + FE		707	706	
	BM + SR + FE		599	601	

4.4 Comparing shift work rules and fatigue scores

An interesting question is how well existing rules and regulations protect different employees from fatigue. To investigate this, we run n instances drawn as described in Section 4.1 for the following cases:

1. The base model with the shift work rules-extension. This case is abbreviated as BM+SR.

2. The base model with the fatigue extension. This case is abbreviated as BM+FE.

3. The base model with both the shift work rules extension and the fatigue extension. This case is abbreviated as BM+SR+FE.

For all three cases we evaluate the solutions to find the real maximum fatigue values. Results are presented in Tables 8. ⇐
K: ***Could perhaps have average scores as the top rows.*** ⇐
K: ***Get data on worst case fatigue in all three cases. Get data on cases of breaking shift work constraints in the base+fatigue case. Get data on nr of shifts worked by each nurse of different bio types. Print some rosters for all nurses or just some interesting bio-profiles.***

4.5 The effects of reduced staff levels on maximum fatigue levels

For each of the n instances used in our tests in Section 4.4, we remove the last nurse to create another instance of 29 nurses, then the second last to create an instance of 28 nurses etc. This way we can analyse how staffing levels affect the worst case fatigue levels. For simplicity, we only include the real objective function values in this run. Results are presented in Table 10

Test for fewer nurses as well. Perhaps it is feasible for BM+FE, although SR make it infeasible. Shifting from SR to FE could then be economically wise.

Table 9: The values of the maximum fatigue values for different cases are presented.

Instance	Nr. of nurses	Real maximum fatigue values for different cases		
		BM+SR	BM+FE	BM+SR+FE
1	30	652	632	659 (656)
	27	656	769	**656**
2	30	662	600	658 (654)
	27	654	661	**654**
3	30	601	706 (660)	601
	27	651	**654**	651

Table 10: The values of the maximum fatigue values for different cases are presented.

Instance	Nr. of nurses	Approx. maximum fatigue values for different cases				
		BM+SR	BM+FE	BM+SR+FE	FRE	avg(Approx-FRE)
1	30	664	590	595	597	5.73
	27	666	599	602	662	6.18
2	30	634	590	595	598	2.13
	27	666	597	597	601	3.19
3	30	599	590	590	604	2.40
	27	666	597	597	602	4.33
4	30	656	595	595	600	3.43
	27	634	597	597	600	2.26

4.6 The value of knowing each individual's biotype

generate roster with alertness model for the default biotype. THen, apply the actual biotypes and see how this affects the model.

Table 11: Effect on the fatigue levels when following roster generated for different biotype (BM + FE).

Instance	Target instance	Original fatigue	Target fatigue
2	1	769	811
1	2		
1	3		

5 Conclusions and future work

Our results demonstrate that current rules and regulations are not able to fully protect nurses from high levels of subjective fatigue. **MARK:** ***Do they?*** This ⇐ implies that the need for fatigue models is large. Furthermore, models should have some way of taking into account individual differences. Explicitly exploring biological details among staff would likely be highly controversial, but more available parameters like age, normal sleep time and chronotype can be used or estimated if nurses agree without controversy.

⇐

MARK: ***I think these are new issues and belong maybe after the experiments. The conclusion should not say anything we didn't have before.***
Although some nurses are more resilient to the fatigue that occurs, they are unlikely to be willing to work extreme schedules to reduce the risk of colleagues who work much fewer night shifts. The decision of how to divide night shifts thus becomes a matter of balancing horizontal and vertical equity in rostering: Is fairness in rostering about treating nurses exactly the same or is it about every nurse contributing to the best of their ability? As traditional shift work rules treat nurses homogeneously, they ensure some level of horizontal equity. The vertical equity is often left unregulated and considered a management issue (Need cite?). This may very well be a reasonable approach. However, the results above suggest that there are great differences in how individuals handle working night shifts.

A way to ensure both the horizontal and vertical equity in rostering if fatigue model is implemented yet could be to base rostering the following rules of thumb:

1. Create a strict set of shift work rules and regulations created by nurses and their trade unions.

2. Introduce a system for insentivising nurses that work night shifts, e.g. extra salary.

3. Explicitly state that shift work rules and regulations can be disregarded on individual levels only if both the nurse and their manager agrees. The nurse must agree to avoid being pushed to take more night shifts than he or her can handle. The manager must also be able to turn down a nurse's

request to work extra night shifts, so as to avoid sickness among staff or over-coverage of insentivised shifts.

Point 1 ensures the horizontal equity, while the combination of points 2 and 3 will insentivise vertical equity. If an individual has so little resilience to shift work that he or she becomes very fatigued despite all rules and regulations, this is problematic, and the individual is likely not fit to perform shift work long-term.
future work: -Create an approximation of circadian rhythm to improve the approximation and make it useful for cases where the circadian rhythm is changed significantly as well. Could use some variant of regression for example. -Test same as improvement occurs in sleep modelling -Introduce fatigue models in real-life applications -Work on check-and-improve process

References

T. Åkerstedt, S. Folkard, and C. Portin. Predictions from the three-process model of alertness. *Aviation, space, and environmental medicine*, 75(3):A75–A83, 2004.

E. I. Ásgeirsson and G. L. Sigurardóttir. Near-optimal MIP solutions for preference based self-scheduling. *Annals of Operations Research*, 239(1):273–293, Apr 2016. ISSN 1572-9338. doi: 10.1007/s10479-014-1597-3. URL https://doi.org/10.1007/s10479-014-1597-3.

M. A. Awadallah, A. L. Bolaji, and M. A. Al-Betar. A hybrid artificial bee colony for a nurse rostering problem. *Applied Soft Computing*, 35:726 – 739, 2015. ISSN 1568-4946. doi: https://doi.org/10.1016/j.asoc.2015.07.004. URL http://www.sciencedirect.com/science/article/pii/S1568494615004342.

G. Beddoe, S. Petrovic, and J. Li. A hybrid metaheuristic case-based reasoning system for nurse rostering. *Journal of Scheduling*, 12(2):99, 2009.

J. Beliën and E. Demeulemeester. On the trade-off between staff-decomposed and activity-decomposed column generation for a staff scheduling problem. *Annals of Operations Research*, 155(1):143–166, 2007.

D. Boivin and P. Boudreau. Impacts of shift work on sleep and circadian rhythms. *Pathologie Biologie*, 62(5):292 – 301, 2014. ISSN 0369-8114. doi: https://doi.org/10.1016/j.patbio.2014.08.001. URL http://www.sciencedirect.com/science/article/pii/S0369811414001230. Sommeil / Sleep.

A. A. Borbély. A two process model of sleep regulation. *Hum neurobiol*, 1(3):195–204, 1982.

Z. E. Bowden. *Behavioral Logistics and Fatigue Management in Vehicle Routing and Scheduling Problems*. PhD thesis, Virginia Tech, 2016.

E. K. Burke, P. De Causmaecker, G. V. Berghe, and H. Van Landeghem. The state of the art of nurse rostering. *Journal of scheduling*, 7(6):441–499, 2004.

E. K. Burke, T. Curtois, G. Post, R. Qu, and B. Veltman. A hybrid heuristic ordering and variable neighbourhood search for the nurse rostering problem. *European journal of operational research*, 188(2):330–341, 2008.

D. Dawson and A. Fletcher. A quantitative model of work-related fatigue: background and definition. *Ergonomics*, 44(2):144–163, 2001. doi: 10.1080/00140130119399. URL https://doi.org/10.1080/00140130119399. PMID: 11209874.

D. Dawson, Y. I. Noy, M. Härmä, T. Åkerstedt, and G. Belenky. Modelling fatigue and the use of fatigue models in work settings. *Accident Analysis & Prevention*, 43(2):549 – 564, 2011. ISSN 0001-4575. doi: https://doi.org/10.

1016/j.aap.2009.12.030. URL `http://www.sciencedirect.com/science/article/pii/S0001457510000059`. Advancing Fatigue and Safety Research.

A. Dohn and A. Mason. Branch-and-price for staff rostering: An efficient implementation using generic programming and nested column generation. *European Journal of Operational Research*, 230(1):157–169, 2013.

N. R. Downing. *Scheduling and rostering with learning constraint solvers*. PhD thesis, 2016.

A. Ernst, H. Jiang, M. Krishnamoorthy, and D. Sier. Staff scheduling and rostering: A review of applications, methods and models. *European Journal of Operational Research*, 153:3–27, 2004.

D. B. Forger, M. E. Jewett, and R. E. Kronauer. A simpler model of the human circadian pacemaker. *Journal of Biological Rhythms*, 14(6):533–538, 1999. doi: 10.1177/074873099129000867. URL `https://doi.org/10.1177/074873099129000867`.

A. Goel and T. Vidal. Hours of service regulations in road freight transport: An optimization-based international assessment. *Transportation Science*, 48 (3):391–412, 2014. doi: 10.1287/trsc.2013.0477. URL `https://doi.org/10.1287/trsc.2013.0477`.

S. Haspeslagh, P. De Causmaecker, A. Schaerf, and M. Stølevik. The first international nurse rostering competition 2010. *Annals of Operations Research*, 218(1):221–236, Jul 2014. ISSN 1572-9338. doi: 10.1007/s10479-012-1062-0. URL `https://doi.org/10.1007/s10479-012-1062-0`.

R. G. Hughes and A. E. Rogers. Are you tired?: Sleep deprivation compromises nurses' health—and jeopardizes patients. *AJN The American Journal of Nursing*, 104(3):36–38, 2004.

S. R. Hursh, D. P. Redmond, M. L. Johnson, D. R. Thorne, G. Belenky, T. J. Balkin, W. F. Storm, J. C. Miller, and D. R. Eddy. Fatigue models for applied research in warfighting. *Aviation, space, and environmental medicine*, 75(3): A44–A53, 2004.

M. Ingre, W. Van Leeuwen, T. Klemets, C. Ullvetter, S. Hough, G. Kecklund, D. Karlsson, and T. Åkerstedt. Validating and extending the three process model of alertness in airline operations. *PLOS ONE*, 9(10):1–15, 10 2014. doi: 10.1371/journal.pone.0108679. URL `https://doi.org/10.1371/journal.pone.0108679`.

R. Jamshidi. Stochastic human fatigue modeling in production systems. *Journal of Industrial and Systems Engineering*, 12(1):270–283, 2019.

R.-C. Lin, M. Y. Sir, E. Sisikoglu, K. Pasupathy, and L. M. Steege. Optimal nurse scheduling based on quantitative models of work-related fatigue. *IIE Transactions on Healthcare Systems Engineering*, 3(1):23–38, 2013.

C. C. Liu and T. C. Wang. Optimal aircraft maintenance crews work shifts with integer programming. In *Applied Mechanics and Materials*, volume 319, pages 479–484. Trans Tech Publ, 2013.

M. Mallis, S. Mejdal, T. T Nguyen, and D. Dinges. Summary of the key features of seven mathematical models of human fatigue and performance. *Aviation, space, and environmental medicine*, 75:A4–14, 04 2004.

P. McCauley, L. Kalachev, A. Smith, G. Belenky, D. Dinges, and H. Van Dongen. A new mathematical model for the homeostatic effects of sleep loss on neurobehavioral performance. *Journal of theoretical biology*, 256:227–239, 2 2009. doi: 10.1016/j.jtbi.2008.09.012. URL `https://doi.org/10.1016/j.jtbi.2008.09.012`.

J.-P. Métivier, P. Boizumault, and S. Loudni. Solving nurse rostering problems using soft global constraints. In *International Conference on Principles and Practice of Constraint Programming*, pages 73–87. Springer, 2009.

H. Meyer auf'm Hofe. Solving rostering tasks as constraint optimization. In E. Burke and W. Erben, editors, *Practice and Theory of Automated Timetabling III*, pages 191–212, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg. ISBN 978-3-540-44629-3.

G. Michalos, S. Makris, L. Rentzos, and G. Chryssolouris. Dynamic job rotation for workload balancing in human based assembly systems. *CIRP Journal of Manufacturing Science and Technology*, 2(3):153 – 160, 2010. ISSN 1755-5817. doi: https://doi.org/10.1016/j.cirpj.2010.03.009. URL `http://www.sciencedirect.com/science/article/pii/S1755581710000167`. Sustainable Development of Manufacturing Systems.

G. Michalos, S. Makris, and G. Chryssolouris. The effect of job rotation during assembly on the quality of final product. *CIRP Journal of Manufacturing Science and Technology*, 6(3):187–197, 2013.

F. Mischek and N. Musliu. Integer programming model extensions for a multi-stage nurse rostering problem. *Annals of Operations Research*, 275(1):123–143, Apr 2019. ISSN 1572-9338. doi: 10.1007/s10479-017-2623-z. URL `https://doi.org/10.1007/s10479-017-2623-z`.

S. Muecke. Effects of rotating night shifts: literature review. *Journal of Advanced Nursing*, 50(4):433–439, 2005. doi: 10.1111/j.1365-2648.2005.03409.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2648.2005.03409.x`.

N. Musliu, J. Gärtner, and W. Slany. Efficient generation of rotating workforce schedules. *Discrete Applied Mathematics*, 118(1-2):85–98, 2002.

N. Nethercote, P. J. Stuckey, R. Becket, S. Brand, G. J. Duck, and G. Tack. Minizinc: Towards a standard cp modelling language. In C. Bessière, editor, *Principles and Practice of Constraint Programming – CP 2007*, pages 529–543, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. ISBN 978-3-540-74970-7.

S. Petrovic and G. V. Berghe. A comparison of two approaches to nurse rostering problems. *Annals of Operations Research*, 194(1):365–384, 2012.

A. Phillips and P. Robinson. A quantitative model of sleep-wake dynamics based on the physiology of the brainstem ascending arousal system. *Journal of Biological Rhythms*, 22(2):167–179, 2007. doi: 10.1177/0748730406297512. URL https://doi.org/10.1177/0748730406297512. PMID: 17440218.

A. J. K. Phillips, P. Y. Chen, and P. A. Robinson. Probing the mechanisms of chronotype using quantitative modeling. *Journal of Biological Rhythms*, 25 (3):217–227, 2010. doi: 10.1177/0748730410369208. URL https://doi.org/10.1177/0748730410369208. PMID: 20484693.

R. Pizarro, G. Rivera, R. Soto, B. Crawford, C. Castro, and E. Monfroy. Constraint-based nurse rostering for the valparaíso clinic center in chile. In *International Conference on Human-Computer Interaction*, pages 448–452. Springer, 2011.

S. Postnova. Sleep modelling across physiological levels. *Clocks & Sleep*, 1(1): 166–184, 2019. ISSN 2624-5175. URL https://www.mdpi.com/2624-5175/1/1/15.

S. Postnova, S. W. Lockley, and P. A. Robinson. Sleep propensity under forced desynchrony in a model of arousal state dynamics. *Journal of Biological Rhythms*, 31(5):498–508, 2016. doi: 10.1177/0748730416658806. URL https://doi.org/10.1177/0748730416658806. PMID: 27432116.

P. Rajdev, D. Thorsley, S. Rajaraman, T. L. Rupp, N. J. Wesensten, T. J. Balkin, and J. Reifman. A unified mathematical model to quantify performance impairment for both chronic sleep restriction and total sleep deprivation. *Journal of Theoretical Biology*, 331:66 – 77, 2013. ISSN 0022-5193. doi: https://doi.org/10.1016/j.jtbi.2013.04.013. URL http://www.sciencedirect.com/science/article/pii/S0022519313001811.

G. Roach, A. Fletcher, and D. Dawson. A model to predict work-related fatigue based on hours of work. *Aviation, space, and environmental medicine*, 75: A61–9; discussion A70, 04 2004.

E. Rönnberg and T. Larsson. Automating the self-scheduling process of nurses in swedish healthcare: a pilot study. *Health Care Management Science*, 13 (1):35–53, Mar 2010. ISSN 1572-9389. doi: 10.1007/s10729-009-9107-x. URL https://doi.org/10.1007/s10729-009-9107-x.

Safe Work Australia. Guide for managing the risk of fatigue at work. https://www.safeworkaustralia.gov.au/system/files/documents/

`1702/managing-the-risk-of-fatigue.pdf`, 2013. Online; accessed 22 August 2019.

L. F. Shampine and M. W. Reichelt. The matlab ode suite. *SIAM journal on scientific computing*, 18(1):1–22, 1997.

M. A. St. Hilaire, M. Rüger, F. Fratelli, J. T. Hull, A. J. K. Phillips, and S. W. Lockley. Modeling Neurocognitive Decline and Recovery During Repeated Cycles of Extended Sleep and Chronic Sleep Deficiency. *Sleep*, 40(1), 12 2016. ISSN 0161-8105. doi: 10.1093/sleep/zsw009. URL `https://doi.org/10.1093/sleep/zsw009`.

A. P. Tvaryanas and N. L. Miller. Human systems integration domain tradeoffs in optimized manning-the task effectiveness scheduling tool. Technical report, NAVAL POSTGRADUATE SCHOOL MONTEREY CA DEPT OF OPERATIONS RESEARCH, 2010.

T.-C. Wang and G.-C. Ke. Fatigue minimization work shift scheduling for air traffic controllers. *International Journal of Automation and Smart Technology*, 3(2):91–99, 2013.

T.-C. Wang and C.-C. Liu. Optimal work shift scheduling with fatigue minimization and day off preferences. *Mathematical Problems in Engineering*, 2014, 2014.

E. Åhsberg, F. Garnberale, and A. Kjellberg. Perceived quality of fatigue during different occupational tasks development of a questionnaire. *International Journal of Industrial Ergonomics*, 20(2):121 – 135, 1997. ISSN 0169-8141. doi: https://doi.org/10.1016/S0169-8141(96)00044-3. URL `http://www.sciencedirect.com/science/article/pii/S0169814196000443`.

E. Åhsberg, G. Kecklund, T. Åkerstedt, and F. Gamberale. Shiftwork and different dimensions of fatigue. *International Journal of Industrial Ergonomics*, 26(4):457 – 465, 2000. ISSN 0169-8141. doi: https://doi.org/10.1016/S0169-8141(00)00007-X. URL `http://www.sciencedirect.com/science/article/pii/S016981410000007X`.